HAL open science

Invariant conjugate analysis for exponential families

Pierre Druilhet, Denys Pommeret

▶ To cite this version:

Pierre Druilhet, Denys Pommeret. Invariant conjugate analysis for exponential families. Bayesian Analysis, 2012, 7 (4), pp.903-916. 10.1214/12-BA731 . hal-00803642

HAL Id: hal-00803642
https://hal.science/hal-00803642

Submitted on 22 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian Analysis (2012) **7**, Number 4, pp. 903–916

Invariant Conjugate Analysis for Exponential Families

Pierre Druilhet* and Denys Pommeret†

Abstract. There are several ways to parameterize a distribution belonging to an exponential family, each one leading to a different Bayesian analysis of the data under standard conjugate priors. To overcome this problem, we propose a new class of conjugate priors which is invariant with respect to *smooth reparameterization*. This class of priors contains the Jeffreys prior as a special case, according to the value of the hyperparameters. Moreover, these conjugate distributions coincide with the posterior distributions resulting from a Jeffreys prior. Then these priors appear naturally when several datasets are analyzed sequentially and when the Jeffreys prior is chosen for the first dataset. We apply our approach to inverse Gaussian models and propose full invariant analyses of three datasets.

Keywords: Bayesian inference, conjugate prior, exponential family, inverse Gaussian distribution, Jeffreys prior, sequential analysis.

1 Introduction

A Bayesian statistical model is made of a parametric statistical model with density $f(x|\theta)$, and a prior on the parameters with density $\pi(\theta)$. In this context, as pointed out by Gelman (2004), a transformation of parameters typically suggests a new family of prior distributions. Other authors emphasized the importance of the parameterization in a Bayesian framework, as for instance Slate (1994) for quadratic natural exponential families, or Palmer (1973) for the inverse Gaussian distribution. More technical reparameterizations occur in Mengersen and Robert (1996) changing mean and variance parameters for estimation purposes (see also Robert and Titterington, 1998, for an application with Markov chains). Several illustrations are proposed in the literature, as in phylogenetics in Zwickl and Holder (2004) where two possible parameterizations are considered for transitions of nucleotides: a transition-transversion rate ratio τ, or a proportion of substitutions that are transitions $\phi = \tau/(2+\tau)$.

A solution to this problem was proposed by Jeffreys (1946). He introduced a prior distribution, whose density is the square root of the determinant of the Fisher information matrix, that is invariant with respect to reparameterization. However, despite this invariance property, the Jeffreys prior is not always recommended: the distribution may be improper and it yields a non-informative distribution.

In this paper we consider the problem of constructing a prior family that is invariant

*Department of Mathematics, Blaise Pascal University, Aubire Cedex, France, pierre.druilhet@math.univ-bpclermont.fr

†Institute of Mathematics of Luminy, Marseille, France, pommeret@univmed.fr

© 2012 International Society for Bayesian Analysis DOI:10.1214/12-BA731

with respect to reparameterization. We restrict our attention to the case where the density $f(x|\theta)$ belongs to an exponential family. In that case the prior distribution $\pi(\theta)$ is often chosen among the family of natural conjugate prior distributions in order to derive a simple expression of the posterior distribution.

We propose a new class of conjugate priors for exponential families based on the Jeffreys measure and named Jeffreys Conjugate Priors (JCPs). This class has the advantage of being invariant with respect to *smooth reparameterization*. When the exponential family is natural quadratic (see Morris, 1983), we show that this approach is equivalent to classical conjugate prior families. From a sequential point of view, if several datasets are successively explored and the Jeffreys prior is chosen for the first dataset, the successive posterior distributions belong to the class of JCPs. It is then a natural class of priors for adaptive procedures, leading to an automatic method of calibration for the hyperparameters. This point is illustrated in our study data.

We study the JCPs associated with the well known inverse Gaussian distributions with one or two unknown parameters. This class of distributions belongs to the exponential family and the problem of reparameteriztion has been emphasized by several authors (see for instance Whitmore, 1979, Palmer, 1973, Chhikara and Folks, 1989, Banerjee and Bhattacharyya, 1979). We will look more closely at this class of distributions, obtaining explicit expressions of JCPs. Two cases will be then considered: the first one is the standard inverse Gaussian distribution with one parameter, which belongs to the cubic natural exponential families. The second one is the general inverse Gaussian distribution with two parameters.

The paper is organized as follows: In Section 2, we introduce the notion of invariant Jeffreys conjugate priors. In Section 3, we focus on inverse Gaussian models. In Section 4, we analyse several datasets, one of them by a sequential procedure.

2 Invariant conjugate priors for exponential families

Consider n i.i.d. random variables $X = (X_1, ..., X_n)$, where the distribution P_θ of X_i belongs to an exponential family (see Barndorff-Nielsen, 1978, for more details). The probability density function of P_θ is given by

$$f(x \,|\, \theta) \quad = \quad \exp\{\theta \cdot t(x) - \varphi(\theta)\}\, k(x), \tag{1}$$

where k is some non-negative function and θ stands for the natural parameter assumed to belong to an open set $\Theta \subset \mathbb{R}^d$. Here, $\theta \cdot t(x)$ denotes the scalar product. The density is considered with respect to some Radon measure (in most cases, the Lebesgue measure for continuous distributions and the counting measure for discrete distributions). Standard conjugate prior distributions $\{\pi_{\alpha,\beta}\}_{\alpha,\beta}$ for θ are defined by their density with respect to the Lebesgue measure by

$$\pi_{\alpha,\beta}(\theta) \propto \exp\{\theta \cdot \alpha - \beta\, \varphi(\theta)\}. \tag{2}$$

Diaconis and Ylvisaker (1979) obtained necessary and sufficient conditions on α and β under which $\pi_{\alpha,\beta}$ is a proper distribution. The main interest of conjugate priors is the following updating formula given the observations $x = (x_1, \cdots, x_n)$:

$$\pi_{\alpha,\beta}(\theta \,|\, x) = \pi_{\alpha+\sum_i t(x_i),n+\beta}(\theta). \tag{3}$$

In the same way, many other families of conjugate priors may be constructed by setting

$$\pi_{\alpha,\beta}(\theta) \propto \exp\{\theta \cdot \alpha - \beta\,\varphi(\theta)\}\ \pi_0(\theta), \tag{4}$$

where π_0 is some non-negative function. More generally, families of conjugate priors may be constructed by considering (2) as the density with respect to some σ-finite measure on Θ instead of the Lebesgue measure. The advantage of such families of priors is that formula (3) initially stated for the standard conjugate priors still holds.

Most often, the usual parameterization of an exponential family is not given by the natural parameter θ, but by $\eta = h(\theta)$ where h is a one-to-one transformation twice continuously differentiable (henceforth abbreviated as *smooth reparameterization*). In the literature, two main families of conjugate prior distributions on η are used: the first one is derived from the family of conjugate priors for θ given by (2) and is defined by

$$\pi_{\alpha,\beta}(\eta) \propto \left|\frac{\mathrm{d}\theta(\eta)}{\mathrm{d}\,\eta}\right| \exp\{\theta(\eta)\cdot\alpha + \beta\ \varphi(\theta(\eta))\}. \tag{5}$$

The second one, known as the standard conjugate prior family, is given by

$$\pi_{\alpha,\beta}(\eta) \propto \exp\{\theta(\eta)\cdot\alpha + \beta\ \varphi(\theta(\eta))\}. \tag{6}$$

Note that Formulas (5) and (6) differ by the Jacobian $|\mathrm{d}\theta(\eta)/\mathrm{d}\,\eta|$. In the general case, these two families of conjugate priors are not identical. However, there are some cases where the two families coincide. For example, in the case of natural exponential families, i.e. $t(x_i) = x_i$ with both x_i and θ belonging to $\mathbb{R}$, which are parameterized by the means of the distributions, $m(\theta) = \mathbb{E}(X_i|\theta)$, Consonni and Veronese (1992) showed that the two families of prior distributions defined by (5) and (6) are identical (up to a reparameterization) if and only if (iff) the exponential family is quadratic, which means that the variance of the distribution is a quadratic polynomial in the mean. Quadratic natural exponential families include the binomial, Poisson, negative-binomial, Gaussian, gamma and hyperbolic distributions (see Morris, 1983). Thus, for these families, we do not take into account the Jacobian for reparameterization. Some extensions to the multivariate case have been established by Gutiérrez-Peña and Smith (1995) (see also Consonni et al. 2004, Gutiérrez-Peña and Rueda, 2003).

We propose a new family of conjugate priors for exponential families that is invariant with respect to reparameterization and that may approach the Jeffreys prior. For any *smooth parameterization* $\eta(\theta)$, we define the corresponding *Jeffreys Conjugate Prior* (JCP) by

$$\pi^J_{\alpha,\beta}(\eta) \propto \exp\{\alpha\cdot\theta(\eta) - \beta\ \varphi(\theta(\eta))\}\ |I_\eta(\eta)|^{1/2}, \tag{7}$$

where $I_\eta(\eta)$ is the Fisher information matrix for η. The special case $\eta = \theta$ leads to

$$\pi^J_{\alpha,\beta}(\theta) \propto \exp\{\theta\cdot\alpha - \beta\,\varphi(\theta)\}\ \ |I_\theta(\theta)|^{\frac{1}{2}}\,, \tag{8}$$

which corresponds to (4) with $\pi_o(\theta) = |I_\theta(\theta)|^{\frac{1}{2}}$, the Jeffreys prior.

It is worth pointing out that, from (3), when the prior distribution is the Jeffreys one, the posterior distribution belongs to the family of JCPs with $\alpha = \sum t(x_i)$ and $\beta = n$. This implies that the JCPs appear naturally in sequential estimation procedures, starting with the Jeffreys prior.

The following proposition establishes the invariance property of JCPs under *smooth reparameterization*.

Proposition 1. *Consider the JCPs class $\{\pi^J_{\alpha,\beta}(\theta)\}_{\alpha,\beta}$ for the parameter θ and let $\eta = h(\theta)$ be a smooth reparameterization. The prior distribution $\pi^J_{\alpha,\beta}(\eta)$ defined by (7) is the same as the distribution derived from $\pi^J_{\alpha,\beta}(\theta)$; that is,*

$$\pi^J_{\alpha,\beta}(\eta) = \left|\frac{\mathrm{d}\theta(\eta)}{\mathrm{d}\,\eta}\right| \pi^J_{\alpha,\beta}(\theta(\eta)).$$

Proof. The proof follows immediately from the invariance property of the Jeffreys measure:

$$|I_\eta(\eta)|^{1/2} = |I_\theta(\theta(\eta))|^{1/2}\left|\frac{\mathrm{d}\theta(\eta)}{\mathrm{d}\,\eta}\right|.$$

■

For a univariate quadratic natural exponential family, the proposition below shows that the JCPs are equivalent to the standard conjugate priors when the parameter is $m(\theta) = \mathbb{E}(X_i|\theta)$.

Proposition 2. *The JCPs associated with the mean parameter are equivalent (up to a reparameterization) to the standard conjugate priors given by (5) iff the univariate natural exponential family is quadratic.*

Proof. Let $V(m) = \mathrm{Var}_m(X_i)$ be the variance function associated with the exponential family. It is well known that $V(m) = |I_m(m)|^{-1}$. Therefore we have to characterize the fact that there exists α_o and β_o such that $V(m) \propto \exp\{\theta(m)\cdot\alpha_o+\beta_o\,\varphi(\theta(m))\}$. Deriving with respect to m this proportionality we get: $V'(m) \propto \theta'(m)\big(\alpha_o+\beta_o\,\varphi'(\theta(m))\big)V(m)$. Since $\theta'(m) = \big(V(m)\big)^{-1}$ and $\varphi'(\theta(m)) = m$, we obtain $V'(m) \propto \alpha_o + \beta_o\, m$; that is, the exponential family is quadratic.

■

As a consequence of Proposition 2, when the model corresponds to a quadratic exponential family, it is not necessary to distinguish between standard conjugate priors and JCPs for the mean parameter. Another advantage of the invariance property stated in Proposition 1 is that conditions on α and β leading to proper prior distributions for $\pi^J_{\alpha,\beta}(\eta)$ do not depend on the choice of the *smooth parameterization* η. In the next section, we shall examine these conditions in the case of inverse Gaussian distributions.

Remark 2.1. *The construction of JCPs can be related to the work of Druilhet and Marin (2007). They proposed to use Jeffreys MAP (JMAP) estimators defined by*

$$JMAP(\eta) = \operatorname*{Argmax}_{\eta} \ \pi(\eta|x) \ |I_\eta(\eta)|^{-\frac{1}{2}}.$$

Unlike MAP estimators, JMAP estimators are invariant with respect to smooth reparameterization. In the case of JCPs, we have:

$$JMAP(\eta) = \operatorname*{Argmax}_{\eta>0} \ \pi^J_{\alpha,\beta}(\eta|x) \ |I_\eta(\eta)|^{-\frac{1}{2}} \ = \ \operatorname*{Argmax}_{\eta>0} \ \pi_{\alpha,\beta}(\eta|x),$$

where the prior $\pi_{\alpha,\beta}(\eta)$ is defined by (6). We may note that JMAP estimators with prior $\pi^J_{\alpha,\beta}(\eta)$ are equivalent to standard MAP estimators with prior $\pi_{\alpha,\beta}(\eta)$, even though both prior distributions are different.

Similarly to the JMAP estimator, Druilhet and Marin (2007) obtained invariant highest posterior density (HPD) based on the Jeffreys measure and called JHPD:

$$JHPD^\gamma(\eta) = \{\eta > 0 : \ \pi^J_{\alpha,\beta}(\eta|x) \ |I_\eta(\eta)|^{-\frac{1}{2}} \geq k_\gamma\} \ = \ \{\eta > 0 : \ \pi_{\alpha,\beta}(\eta) \geq k_\gamma\},$$

for some constant k_γ.

3 Invariant conjugate analysis for inverse Gaussian distributions

The inverse Gaussian distribution appears in many probabilistic models and has a wide range of applications (see Seshadri, 1993, for more details). Depending on the context, several parameterizations have been proposed. For instance Tweedie (1956) proposed the following parameterization for the density function:

$$f(x;\mu,\lambda) \quad = \quad \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left\{\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right\}, \quad x > 0,$$

where $\mu > 0$ denotes the mean parameter and $\lambda > 0$ stands for the shape parameter. The variance μ^3/λ is sometimes called the *Jorgensen parameter* in the literature of dispersion models (see Jorgensen, 1986). If $X_1, \cdots, X_n$ are i.i.d. random variables with density $f(x;\mu,\lambda)$, then $\sum X_i$ has density $f(x;n\mu,n^2\lambda)$. The first time level 1 is attained for a real Brownian motion process with drift $\psi = 1/\mu$ and diffusion constant $1/\lambda$ has inverse Gaussian distribution (see for instance Cox and Miller, 1965). The parameter $\phi = \lambda/\mu$ determines the shape of the distribution, and, for λ fixed, as ϕ increases the inverse Gaussian tends to the normal distribution (see Seshadri, 1993, for some illustrations). The parameter $\delta = (\phi\mu)^{-1/2}$ is the coefficient of variation. In the particular case where $\lambda = 1 = \phi$, we obtain the so called *standard inverse Gaussian distribution.*

In this context, several problems of parameterization appeared in the literature. For example, with the parameterization $(\psi = 1/\mu, \lambda)$, Whitmore (1979) obtained a normal-gamma family of conjugate priors. Palmer (1973) showed that the natural conjugate prior does not exist for the parameterization (μ, λ), but it exists for λ when μ is known. Chhikara and Folks (1989) showed that the natural conjugate prior exists for the parameter $\psi = 1/\mu$. With the parameterization $(\psi = 1/\mu, \lambda)$ Barnejee and Bhattacharyya (1979) proved that the Jeffreys prior (Jeffreys, 1961) yields an improper posterior distribution for ψ.

We proceed with the study of the JCPs associated with the inverse Gaussian distribution with one or two parameters. We consider an i.i.d. sample $x = (x_1, ..., x_n)$ from an inverse Gaussian distribution and we denote $S = \sum_{i=1}^n x_i$ and $T = \sum_{i=1}^n 1/x_i$.

3.1 Standard inverse Gaussian distribution

We set $\lambda = 1$ and consider the mean parameter $\mu > 0$. The density of the standard inverse Gaussian distribution is given by

$$f(x;\mu) \quad = \quad \left(\frac{1}{2\pi x^3}\right)^{1/2} \exp\left\{\frac{-(x-\mu)^2}{2\mu^2 x_i}\right\}, \ x > 0,$$

with respect to Lebesgue measure on $\mathbb{R}^+$. The Fisher information for μ is $I_\mu(\mu) = \mu^{-3}$ and the JCP is given by

$$\pi^J_{\alpha,\beta}(\mu) \propto \mu^{-3/2} \exp\left\{\frac{-\alpha}{2\mu^2} + \beta/\mu\right\} \mathbb{I}_{(0,\infty)}(\mu),$$

where $\mathbb{I}_{(0,\infty)}$ denotes the indicator function on $(0, \infty)$. For $\alpha = 0$, $\pi^J_{\alpha,\beta}$ is an inverse gamma distribution with parameters $(1/2, -\beta)$. It is proper iff $\beta < 0$. Its mode is $-2\beta/3$. For $\alpha > 0$, $\pi^J_{\alpha,\beta}$ is proper with mode

$$M \quad = \quad \frac{1}{3}\left\{\sqrt{\beta^2 + 6\alpha} - \beta\right\}.$$

In both cases, the means of the prior and posterior distributions are $+\infty$. Note that $\pi^J_{\alpha,\beta}$ corresponds to a left-truncated generalized inverse normal distribution (see Robert, 1991) defined by

$$f_{\sigma^2,\xi}(m) = |m|^{-c} \exp\left\{\frac{-1}{2\sigma^2 m^2} + \frac{\xi}{\sigma^2 m}\right\},$$

with $c = 3/2$, $\sigma^2 = 1/\alpha$ and $\xi/\sigma^2 = \beta$.

Using the updating formula (3), the posterior distribution is given by

$$\pi^J_{\alpha,\beta}(\mu \,|\, x) = \pi^J_{\alpha+S,n+\beta}(\mu).$$

Since the posterior mean does not exist, we use the MAP estimator for μ:

$$\mathrm{MAP}(\mu) \quad = \quad \frac{1}{3}\left\{\sqrt{(\beta+n)^2 + 6(\alpha+S)} - (\beta+n)\right\},$$

which is not invariant with respect to reparameterization. If $\beta + n > 0$, an invariant estimator is

$$\mathrm{JMAP}(\mu) = \frac{\alpha + S}{\beta + n}.$$

If $\beta + n \leq 0$, then $\mathrm{JMAP}(\mu) = +\infty$.

3.2 Inverse Gaussian model with two parameters

We consider the parameterization $(\psi = 1/\mu, \lambda)$, $\psi > 0$ and $\lambda > 0$. The determinant of the Fisher information matrix is given by

$$|I(\psi, \lambda)| \quad = \quad (2\psi\lambda)^{-1},$$

and the density functions of the JCPs with respect to the Lebesgue measure on $\mathbb{R}^+ \times \mathbb{R}^+$ are given by

$$\pi^J_{\alpha,\beta}(\psi, \lambda) \quad \propto \quad \exp\left\{-\frac{\lambda}{2}\left(\alpha_1\psi^2 - 2\beta\psi + \alpha_2\right)\right\}\psi^{-1/2}\lambda^{(\beta-1)/2}. \tag{9}$$

Note that, conditioning on λ, we get

$$\pi^J_{\alpha,\beta}(\psi|\lambda) \propto \exp\{\alpha_1\psi^2 + \alpha_2\psi\}\ \psi^{-1/2},$$

and conditioning on ψ, $\pi^J_{\alpha,\beta}(\lambda|\psi)$ is a gamma distribution.

For natural exponential families, when the density function of the conjugate prior is given by $\exp\{\alpha.\theta - \beta\,\varphi(\theta)\}$ with respect to the Lebesgue measure, Diaconis and Ylvisaker (1979) proved that the prior is proper for $\beta > 0$ and α/β in the interior of the convex hull of the support of a reference measure γ satisfying $\exp\{\varphi(\theta)\} = \int \exp\{\theta x\}\gamma(dx)$. In the case of inverse Gaussian models, the following result provides a criterion for JCPs to be proper.

Proposition 3. *i) The JCP distribution (9) is proper iff $\alpha_1 > 0$, $\alpha_2 > 0$ and $-1/2 \leq \beta < \sqrt{\alpha_1\alpha_2}$.*

ii) $\mathbb{E}(\lambda) < \infty$ iff $\alpha_1 > 0$, $\alpha_2 > 0$ and $-5/2 \leq \beta < \sqrt{\alpha_1\alpha_2}$.

iii) $\mathbb{E}(\psi) < \infty$ iff $\alpha_1 > 0$, $\alpha_2 > 0$ and $1/2 \leq \beta < \sqrt{\alpha_1\alpha_2}$.

Proof. For any given $\psi > 0$, $\pi^J_{\alpha,\beta}(\lambda|\psi)$ is a gamma distribution. It is integrable with respect to $0 < \lambda < +\infty$ iff $\beta > -1$ and $\alpha_1\,\psi^2 - 2\beta\psi + \alpha_2 > 0$. In these cases, the marginal distribution is

$$\pi^J_{\alpha,\beta}(\psi) \quad \propto \quad \psi^{-1/2}(\alpha_1\psi^2 - 2\beta\psi + \alpha_2)^{-\frac{\beta+1}{2}}.$$

Then, it is straightforward to check that the conditions for integrability with respect to λ and ψ are equivalent to those stated in *i*). We can prove *ii*) and *iii*) with the same argument.

■

MAP estimators for (ψ, λ) cannot be used here. Indeed, the posterior distribution is equal to $+\infty$ for $\psi = 0$ whatever the data are; consequently, the MAP estimator of λ is undefined. This is one more reason to use JMAP estimators. For $\pi^J_{\alpha,\beta}(\psi, \lambda)$ as the prior, we have

$$\begin{aligned} \text{JMAP}(\psi) &= \frac{\beta + n}{(\alpha_1 + S)}, & (10) \\ \text{JMAP}(\lambda) &= \frac{(\alpha_1 + S)(\beta + n)}{(\alpha_1 + S)\,(\alpha_2 + T) - (\beta + n)^2}. & (11) \end{aligned}$$

For new parameters, the posterior analysis using JCP and JMAP is invariant provided the Fisher information is well defined. Therefore, all the results concerning the new parameters can be easily obtained from the previous ones. Consider, for example, the parameters $(\mu = \frac{1}{\psi}, \lambda)$. The Jeffreys prior is $\mu^{-\frac{3}{2}}\,\lambda^{-\frac{1}{2}}$ and the JCP is

$$\pi^J_{\alpha,\beta}(\mu, \lambda) \quad \propto \quad \exp\left\{-\frac{\lambda}{2}\left(\frac{\alpha_1}{\mu^2} - \frac{2\beta}{\mu} + \alpha_2\right)\right\}\mu^{-3/2}\lambda^{(\beta-1)/2}.$$

Since JCPs are invariant, the conditions on α, β for $\pi^J_{\alpha,\beta}(\mu, \lambda)$ to be proper are identical to those stated in Proposition 3. It is worth noting that, by invariance, the JMAP estimator of λ is not affected by the change of ψ into μ and the JMAP estimator of μ satisfies $\text{JMAP}(\mu) = (\text{JMAP}(\psi))^{-1}$.

The improper non-informative Jeffreys prior, denoted by π^J, corresponds to $\alpha_1 = \alpha_2 = \beta = 0$ whatever the parameterization is. Therefore the JCPs family appears to be a convenient way to approximate π^J by an invariant class of priors. We just have to choose α_1, α_2 and β close to 0 with $\beta^2 < \alpha_1\alpha_2$. The posterior distribution is given by

$$\pi^J_{\alpha,\beta}((\psi, \lambda)|x) = \pi^J_{\alpha_1+S,\alpha_2+T,\beta+n}(\psi, \lambda). \tag{12}$$

By invariance, we have a similar formula for the parameter (μ, λ):

$$\pi^J_{\alpha,\beta}((\mu, \lambda)|x) = \pi^J_{\alpha_1+S,\alpha_2+T,\beta+n}(\mu, \lambda).$$

Corollary 4. *i) The Jeffreys prior yields an improper posterior distribution for $n = 1$ and a proper posterior distribution for $n > 1$.*

ii) For n large enough, any JCP leads to a proper posterior distribution.

Proof. *i)* It is sufficient to show that the conditions stated in Proposition 3 are satisfied with $\alpha_1 = S$, $\alpha_2 = T$ and $\beta = n$. By Jensen's inequality, $S\,T = \sum_{i=1}^n x_i \sum_{i=1}^n \frac{1}{x_i} \geq n^2$. Equality holds iff all the x_i are equal which occurs with probability 0 for $n > 1$ and

probability 1 for $n = 1$. The result follows.
ii) It is sufficient to prove that for large values of n, $(\alpha_1 + S)/n > 0$, $(\alpha_2 + T)/n > 0$ and $(\alpha_1/n + S/n)(\alpha_2/n + T/n) > (\beta/n + 1)^2$. By the law of large numbers, for any given $\lambda > 0$ and $\mu > 0$, the almost sure limits of $(\alpha_1 + S)/n$ and $(\alpha_2 + T)/n$ when n tends to $+\infty$ are respectively μ and $1/\lambda + 1/\mu$. The result follows for n large enough.

■

For the Jeffreys prior, the JMAP estimator is equivalent to the Maximum Likelihood estimator (see Druilhet and Marin, 2007). Fixing $\beta = \alpha_1 = \alpha_2 = 0$ in (10) and (11), we obtain

$$\widehat{\psi}_{\mathrm{ML}} = \frac{n}{S}, \quad \text{and} \quad \widehat{\lambda}_{\mathrm{ML}} = \frac{n\,S}{S\,T - n^2}. \tag{13}$$

The JHPD for the parameter ψ is the set $\{\psi > 0 :\ \alpha_1\psi^2 - 2\beta\psi + \alpha_2 - k_\gamma^{(\beta+1)/2} \leq 0\}$.

4 Illustrations

In this Section we consider three datasets. The first two are used to assess the sensitivity of the JCP with respect to the choice of the hyperparameters. The third one is used to illustrate the JCPs in a sequential procedure.

4.1 Bayesian estimation with JCPs

We assume that the observations in each dataset are i.i.d. inverse Gaussian with parameters (ψ, λ). The posterior density is given by (12). The case $\alpha_1 = \alpha_2 = \beta = 0$ coincides with the Jeffreys prior. Moreover, by Proposition 3, $\mathbb{E}(\lambda)$ (resp. $\mathbb{E}(\psi)$) exists iff $\alpha_1 > 0$, $\alpha_2 > 0$, $\alpha_1\alpha_2 > \beta^2$ and $\beta > -5/2$ (resp. $\beta > 1/2$). In our numerical studies we choose $\alpha_1 = \alpha_2 = \beta(1+\epsilon)$, with $\epsilon > 0$ and we let β vary from 0 (the Jeffreys prior) to 1000. This set of values allows us to compare the JCP ($\beta > 0$) to the classical Jeffreys prior ($\beta = 0$).

Example 1 The inverse Gaussian distribution was used in Folks and Chhikara (1978) to model a set of data giving runoff amounts at Jug Bridge, Maryland. The data are presented in Table 1.

0.17	0.19	0.23	0.33	0.39	0.39	0.40	0.45	0.52	0.56	0.59	0.64
0.66	0.70	0.76	0.78	0.95	0.97	1.02	1.12	1.24	1.59	1.74	2.92

Table 1: *Runoff amounts at Jug Bridge*

Let us assume a two parameters inverse Gaussian distribution with a Jeffreys conjugate prior given by (9). According to the previous conditions, we fix $\epsilon = 0.1$ and $\alpha_1 = \alpha_2 = \beta(1+\epsilon)$. We let $\beta = 0, 0.01, 0.1, 1, 10, 100, 1000$. Small values of β correspond to the proximity of the Jeffreys prior. For the limit case $\beta = 0$, the JMAP coincides with the maximum likelihood estimator and the JCP is equivalent to the Jeffreys prior. The maximum likelihood estimates of λ and ψ are

$$\widehat{\lambda} = 0.0433, \qquad \widehat{\psi} = 1.3278.$$

JMAPs and posterior expectations, obtained from (10) and (11), are summarized in Table 2. It can be observed that JMAPs and posterior expectations are not too sensitive to the values of the hyperparameters when β is close to zero. For large values of β, it can be seen that these estimates deviate from the values of $\widehat{\lambda}$ and $\widehat{\psi}$. The prior means $\mathbb{E}(\lambda)$ and $\mathbb{E}(\psi)$ are more sensitive to the choice of β and turn out to be unstable when β is close to zero. As noticed by a referee, the prior means seem to be stabilized when β increases.

	$\mathbb{E}(\lambda\|X)$	$\mathbb{E}(\lambda)$	JMAP(λ)	$\mathbb{E}(\psi\|X)$	$\mathbb{E}(\psi)$	JMAP(ψ)
$\beta = 0$	0.044	$+\infty$	0.0433	1.154	$+\infty$	1.328
$\beta = 0.01$	0.044	197.77	0.0434	1.154	$+\infty$	1.327
$\beta = 0.1$	0.044	23.16	0.0435	1.151	$+\infty$	1.325
$\beta = 1$	0.046	5.909	0.0452	1.124	1	1.303
$\beta = 10$	0.062	5.226	0.0618	0.974	0.896	1.165
$\beta = 100$	0.223	5.238	0.2228	0.946	0.908	0.966
$\beta = 1000$	1.413	2.080	1.413	0.915	0.499	0.916

Table 2: JMAP, prior and posterior expectations for *Jug Bridge data*

Example 2 Chhikara and Folks (1977) used the inverse Gaussian distribution to model active repair times (in hours) of an airbone communication transceiver. Table 3 presents these data.

0.2	0.3	0.5	0.5	0.5	0.5	0.6	0.6	0.7	0.7	0.7	0.8	0.8	1.0	1.0	1.0
1.0	1.1	1.1	1.3	1.5	1.5	1.5	1.5	2.0	2.0	2.2	2.5	2.7	3.0	3. 0	3.3
3.3	4.0	4.0	4.5	4.7	5.0	5.4	5.4	7.0	7.5	8.8	9.0	10.3	22.0	24.5	

Table 3: *Active repair times (in hours) of an airbone communication*

Hyperparameters α_1, α_2 and β are chosen as in the previous example. Table 4 contains JMAP and posterior estimates for λ and ψ. The maximum likelihood estimates of ψ and λ are

$$\widehat{\lambda} = 1.6588 \quad , \quad \widehat{\psi} = 0.2772.$$

As previously, it can be seen that for large values of β, posterior and JMAP estimates are different from those obtained by maximum likelihood. Consequently we may suggest small values for β yielding a proper posterior which is close to the vague prior and such that prior means for ψ and λ may coincide with expert judgments.

	$\mathbb{E}(\lambda\|X)$	$\mathbb{E}(\lambda)$	JMAP(λ)	$\mathbb{E}(\psi\|X)$	$\mathbb{E}(\psi)$	JMAP(ψ)
$\beta = 0$	1.657	$+\infty$	1.659	0.270	$+\infty$	0.277
$\beta = 0.01$	1.657	197.77	1.659	0.270	$+\infty$	0.277
$\beta = 0.1$	1.657	23.16	1.659	0.270	$+\infty$	0.278
$\beta = 1$	1.656	5.909	1.657	0.274	1	0.281
$\beta = 10$	1.658	5.226	1.659	0.311	0.896	0.316
$\beta = 100$	1.993	5.238	1.994	0.527	0.908	0.529
$\beta = 1000$	3.787	5.238	3.787	0.826	0.909	0.826

Table 4: JMAP, prior and posterior expectations for *repair data*

Remark 4.1 (Influence of the parameter $\boldsymbol{\epsilon}$.). *The previous estimates were obtained for* $\epsilon = 0.1$. *To evaluate the influence of this parameter on the relation* $\alpha_1 = \alpha_2 = (1+\epsilon)\beta$, *we propose two other values for* ϵ, 0.01 *and* 1, *and compare the results obtained for* $\beta = 0.01$, $\beta = 1$ *and* $\beta = 10$ *on the repair dataset. They are reported in Table 5. JMAP and posterior estimators appear to be robust with respect to* ϵ. *For small values of* β, *both posterior means and JMAP are stable, while prior means may diverge. Here again we may recommend small values of* ϵ, *combined with small values of* β, *to get a proper prior which is close to the vague prior of Jeffreys.*

ϵ	β	$\mathbb{E}(\lambda\|X)$	$\mathbb{E}(\lambda)$	JMAP(λ)	$\mathbb{E}(\psi\|X)$	$\mathbb{E}(\psi)$	JMAP(ψ)
0.01	0.01	1.657	1295.33	1.658	0.270	$+\infty$	0.277
	1	1.661	50.99	1.663	0.274	1	0.281
	10	1.708	0.180	1.709	0.312	102.33	0.318
0.1	0.01	1.657	197.77	1.658	0.270	$+\infty$	0.277
	1	1.656	5.909	1.657	0.274	1	0.281
	10	1.658	5.226	1.659	0.311	0.896	0.316
1	0.01	1.656	36.057	1.658	0.270	$+\infty$	0.277
	1	1.601	1	1.602	0.272	1	0.279
	10	1.283	0.668	1.284	0.293	0.425	0.301

Table 5: Influence of the parameter ϵ for *repair data*

4.2 Sequential procedure

If several samples are observed successively, one can start with the Jeffreys prior for the first dataset. The posterior distribution, which is also the prior distribution for the second data set, is a JCP. Therefore, by (3), the posterior/prior distributions for the next datasets are also JCPs. To illustrate this approach, we consider two datasets from Jorgensen (1980) which consist of the number of operating hours between successive failures of airconditioning equipment in aircrafts. The sample sizes are respectively $n_1 = 23$ and $n_2 = 6$. Table 6 describes these data.

Aircraft 1	413 14 58 37 100 65 9 169 447 184 36 201 118 34 31 18 18 67 57 62 7 22 34
Aircraft 2	194 15 41 29 33 181

Table 6: Number of operating hours between successive failures of airconditioning equipment in 2 aircrafts

We first used the Jeffreys prior on the first failures data. The posterior distribution, say JCP_1, is the JCP with parameters $\beta = n_1$, $\alpha_1 = S_1$ and $\alpha_2 = T_1$, where $S_1 = \sum_{i=1}^{n_1} x_i$ and $T_1 = \sum_{i=1}^{n_1} 1/x_i$, and the x_i, $i = 1, ..., n_1$, stand for the first dataset. We obtained

$$\mathrm{JCP}_1 \quad \propto \quad \exp\left\{-\frac{\lambda}{2}\left(2201\psi^2 - 46\psi + 0.7485\right)\right\}\psi^{-1/2}\lambda^{11}.$$

For the second dataset, we use JCP_1 as the prior distribution. The estimates of ψ and λ based on the Jeffreys posterior are displayed in Table 7. It can be seen that the estimates obtained with JCP_1 are different from those obtained with the Jeffreys prior. Their values heavily rely on the information contained in the first *aircraft data*. For comparison, the estimates obtained for the first dataset with Jeffreys prior were $\mathbb{E}(\lambda|$ Aircraft 1$) = 45.02$ and $\mathbb{E}(\psi|$ Aircraft 1$) = 0.010$.

	JCP_1 (from Aircraft 1)	Jeffreys
$\mathbb{E}(\lambda\|$ Aircraft 2$)$	63.838	47.963
$\mathrm{JMAP}(\lambda)$	64.171	48.102
$\mathbb{E}(\psi\|$ Aircraft 2$)$	0.0106	0.0103
$\mathrm{JMAP}(\psi)$	0.0121	0.0107

Table 7: Influence of the prior for the second *aircraft data*. JCP_1 denotes the prior obtained using the posterior from the first dataset.

This procedure is particularly suitable here because the second aircraft dataset is small. But it would be interesting to take into account the sizes of the successive datasets in the prior. Then, an adaptive prior depending on the sample sizes n_1 and n_2 should be a mixture of the prior JCP_1 and the Jeffreys prior π^J, as follows:

$$\mathrm{Jmix} \quad = \quad p_1\, \mathrm{JCP}_1 + p_2\, \pi^J,$$

where $p_1 = n_1/(n_1+n_2)$ and $p_2 = 1-p_1$. Using the Jmix prior and applying this mixture of priors to the *aircraft dataset* leads to $\mathbb{E}(\lambda|$ Aircraft 2$) = 60.504$ and $\mathbb{E}(\psi|$ Aircraft 2$) = 0.0105$. Here the results obtained for JCP_1 and Jmix are very close since $p_1 = 0.79$.

References

Banerjee, A.K. and Bhattacharyya, G.K. (1979). "Bayesian Results for the Inverse Gaussian Distribution with an Application". *Technometrics*, 21, 247–251.

Barndorff-Nielsen, O.E. (1978). *Information and exponential families in statistical theory.* Wiley, New-York.

Chhikara, R.S. and Folks, J.L. (1977). "The inverse Gaussian distribution as a lifetime model". *Technometrics*, 4, 461–468

Chhikara, R.S. and Folks, J.L. (1989). *The Inverse Gaussian Distribution: Theory, Methodology and Applications.* Marcel Dekker, Inc. New York and Basel.

Consonni, G. and Veronese, P. (1992). "Conjugate priors for exponential families having quadratic variance functions". *Journal of the American Statistical Association*, 87, 1123–1127.

Consonni, G., Veronese, P. and Gutiérrez-Peña, E. (2004). "Reference priors for exponential families with simple quadratic variance function". *Journal of Multivariate Analysis*, 88, 335–364.

Cox, D.R. and Miller, H.D. (1965). *The Theory of Stochastic Processes.* London: Methuen.

Diaconis, P. and Ylvisaker, D. (1979). "Conjugate priors for exponential families". *The Annals of Statistics*, 7, 269–281.

Druilhet, P. and Marin, J.M. (2007). "Invariant HPD credible sets and MAP estimators". *Bayesian Analysis*, 4, 681–692.

Folks, J.L. and Chhikara, R.S. (1978). "The Inverse Gaussian Distribution and its Statistical Application - A Review". *Journal of the Royal Statistical Society B*, 40, 263–289.

Gelman, A. (2004). "Parameterization and Bayesian Modeling". *Journal of the American Statistical Association*, 99, 537–545.

Gutiérrez-Peña, E. and Rueda, R. (2003). "Reference Priors for Exponential Families". *Journal of Statistical Planning and Inference*, 110, 35–54.

Gutiérrez-Peña, E. and Smith, A.F.M. (1995). "Conjugate parametrizations for natural exponential families". *Journal of the American Statistical Association*, 90, 1347–1356.

Jeffreys, H. (1946). "An invariant form for the prior probability in estimation problems". *Proceedings of the Royal Society of London, Series A*, 186, 453–461.

Jeffreys, H. (1961). *Theory of Probability* (3rd Edition. ed.). University Press, Oxford.

Jorgensen, B. (1980). "Statistical properties of the generalized inverse Gaussian distribution". *Memoirs of University of Aarhus*, 4.

Jorgensen, B. (1986). “Some properties of exponential dispersion models”. *Scandinavian Journal of Statistics*, 13, 187–197.

Mengersen, K. and Robert, C.P. (1996). “Testing for mixtures: a Bayesian entropic approach”. In J. O. Berger, J. M. Bernardo, A. P. Dawid, D. V. Lindley and A. F. M. Smith (eds), *Bayesian Statistics*, Vol. 5, 255–76, Oxford University Press.

Morris, C.N. (1983). “Natural exponential families with quadratic variance functions: statistical theory”. *The Annals of Statistics*, 11, 515–529.

Palmer, J. (1973). “Bayesian analysis of the inverse Gaussian distribution”. *Ph.-D. thesis*. Oklahoma State University.

Robert, C.P. (1991). “Generalized inverse normal distribution”. *Statistics & Probability Letters*, 11, 37–41.

Robert, C.P. and Titterington, D.M. (1998). “Reparameterization strategies for hidden Markov models and Bayesian approaches to maximum likelihood estimation”. *Statistics and Computing*, 8, 145–158.

Seshadri, V. (1993). *The inverse Gaussian distribution: a case study in exponential families*. Oxford University Press, New York.

Slate, E.H. (1994). “Parameterizations for Natural Exponential Families with Quadratic Variance Functions”. *Journal of the American Statistical Association*, 89, 1471–1482.

Tweedie, M.C.K. (1956). “Some statistical properties of inverse Gaussian distributions”. *Virginia Journal of Sciences (N.S.)*, 7, 160–165.

Whitmore, G. A. (1979). “An Inverse Gaussian Model for Labour Turnover”. *Journal of the Royal Statistical Society (Serie A)*, 142, 468–478.

Zwickl, D.J. and Holder, M.T. (2004). “Model Parameterization, Prior Distributions, and the General Time-Reversible Model in Bayesian Phylogenetics” *Systematic Biology*, 53, 877–888.

Acknowledgements The authors are grateful to the Editor, the Associate Editor and two anonymous reviewers for their thoughtful comments. We especially thank a reviewer who corrected Proposition 3 and suggested Corollary 4.